From calls to communities: a model for time varying social networks

Guillaume Laurent,[1,2] Jari Saramäki,[2] and Márton Karsai[1,*]

[1]*Laboratoire de l'Informatique du Parallélisme, INRIA-UMR 5668, IXXI, ENS de Lyon, 69364 Lyon, France*
[2]*BECS, Aalto University School of Science, P.O. Box 12200, FI-00076, Finland*
(Dated: June 2, 2015)

Social interactions vary in time and appear to be driven by intrinsic mechanisms, which in turn shape the emerging structure of the social network. Large-scale empirical observations of social interaction structure have become possible only recently, and modelling their dynamics is an actual challenge. Here we propose a temporal network model which builds on the framework of *activity-driven time-varying networks* with memory. The model also integrates key mechanisms that drive the formation of social ties – *social reinforcement*, *focal closure* and *cyclic closure*, which have been shown to give rise to community structure and the global connectedness of the network. We compare the proposed model with a real-world time-varying network of mobile phone communication and show that they share several characteristics from heterogeneous degrees and weights to rich community structure. Further, the strong and weak ties that emerge from the model follow similar weight-topology correlations as real-world social networks, including the role of weak ties.

I. INTRODUCTION

The emergent structure and dynamics of social interactions are consequences of individual and collective decision-making processes [1]. Understanding the driving forces behind social network formation has been a great challenge that has attracted lots of attention during the last decades. Even though conventional experimental approaches (based on e.g. surveys) have revealed fundamental rules of social behaviour [2], it has not been possible scale such approaches up to sizes required for observing features of large-scale social structures, let alone their detailed dynamics. However, technological advances have helped to overcome these limitations through detailed digital recording of social interactions [3, 4]. Following these advancements, a lot of effort has been put on mapping social structures from digital traces, and on establishing methodology for the analysis of social networks connecting millions of individuals. These efforts have led to discover the small-world architecture [5], the heterogeneous network structures induced by preferential attachment mechanisms [6, 7], the different roles of strong and weak ties [8, 9] and their relationship to triangle formation [10] and the emergence of communities [11, 12].

The earliest data-driven studies of social networks were mainly focused on static network structure. However, it is evident that this approach both unnecessarily discards available information – electronic interaction traces typically come with time stamps – as well as hinders detailed understanding of the mechanisms shaping network structure. In the recent years, steps have been taken to come around these limitations, with increased focus on both network dynamics (changes in link structure, such as in [13]) and temporal network structure [14] (time-stamped interaction events between individuals giving rise to *e.g.* temporal motifs [15]). In particular, stud-

ies of temporal networks of human interaction have revealed tie dynamics with memory and reinforcement processes [16, 17], bursty dynamical patterns [18, 19], temporal homophily in communication motifs [20] as well as different strategies of communication and tie maintenance [21].

Modelling temporal networks and their co-evolution with processes unfolding on the networks is still one of the key challenges in the field. Empirical studies employing null models with shuffled interaction sequences [22–24] have pointed out that bursty interactions together with features such as weight-topology correlations strongly influence the speed of any type of information diffusing on the temporal network [22, 24–27]. However, a limiting factor of this approach is that it only allows to remove existing temporal inhomogeneities and correlations from empirical sequences in order to understand their effects. Because of this, the null model approach does not provide ways to continuously adjust the level of inhomogeneities or temporal correlations. This limitation can be overcome by resorting to artificial models of temporal networks [28, 29] that allow to fine-tune the level of different features one by one. One promising direction is the *activity-driven framework* [29] that assumes heterogeneous activity of agents. It provides a general foundation, which can be extended by incorporating additional temporal and structural correlations. Studies of the *activity-driven time-varying network model* have demonstrated that memory effects can explain the emergence of strong and weak ties in social networks [17], while endo- and exogenous tie formation processes may control the evolution of business alliance networks [30]. The same model has been used to study how time-varying interactions influence the dynamics of random walkers [31, 32], rumour spreading [17], and epidemic processes [33].

In this paper, our aim is to consider further social mechanisms that would make models of time-varying social networks more realistic. We introduce a temporal network model with adjustable community structure and emergent weight-topological correlations via the exten-

sion of the activity-driven time-varying network model with three additional mechanisms. These are i) reinforcement processes to model memory-driven interaction dynamics of individuals [17], ii) focal and cyclic closure motivated by the work of Kumpula et. al. [12, 34] to capture patterns responsible for the emerging community structure, and iii) a node removal process. Using this temporal network model we demonstrate the effect of the scalable community structure and social reinforcement on information spreading, which co-evolves with the time-varying interactions. After the introduction of the model, we discuss its temporal behaviour, and match the emerging network properties with empirical observations. We discuss the present higher order correlations and their effects on spreading processes, and finally we conclude our work and discuss possible future directions.

II. MODEL

Our model definition builds on the activity-driven time-varying network (ADN) model introduced by Perra *et al.* [29]. The original model takes N nodes that are each assigned *a priori* with an activity probability per unit time $a_i = \eta x_i$. Here x_i is the activity potential of node i, drawn from a suitable distribution $F(x_i)$ ($x_i \in [\epsilon, 1]$ bounded by a minimum activity ϵ), and η is a time rescaling factor determining the average number of active nodes per unit time to be $\eta \langle x \rangle N$. At the beginning of each time step (time t), the model network $G_t = (V_t, E_t)$ is initially disconnected ($E_t = \emptyset$) and it evolves via the following steps: (i) Through random sequential update each node i becomes active with probability $a_i \Delta t$ and connects to m other randomly selected nodes in G_t. (ii) At the end of each step (time $t + \Delta t$), all created links are removed and the process moves on to the next iteration. Note that during a single step, a node can participate in temporal interactions by creating a link (if it is active), or by receiving link(s) from other active nodes. The activity potentials of humans are commonly observed to be heterogeneous [35–37]. We approximate this with a power-law activity potential distribution $F(x_i) \sim x_i^{-\gamma}$ with exponent $\gamma = 2.8$, based on empirical observations and earlier modelling work [17]. Furthermore, since we aim to model one-to-one interactions, typical for mobile telephone communication, we set $m = 1$. In addition, without loss of generality, we fix the parameters $\eta = 1$, $\epsilon = 10^{-3}$ and $\Delta t = 1$ [17].

A. Social mechanisms

The first mechanism to be integrated with the activity-driven model is *memory* in terms of frequently repeated interactions between a node and its peers who have been contacted before. This has been considered earlier by Karsai *et al.* [17] with *reinforcement processes*, where a node remembers its connected neighbours, and depend-

ing on its degree n at t, either creates a new link with probability $p(n) = c/(n + c)$, or interacts with a neighbour with probability $1 - p(n) = n/(n + c)$ to reinforce an existing tie. The introduction of memory defines a non-Markovian dynamics where past interactions influence the actual decision of a node. The larger a node's egocentric network size n, the smaller the chance for it to create a new link in the actual iteration. This captures the realistic assumption that interactions in a social network are not random but favoured with actual friends whose number is strongly limited by cognitive capacities [38, 39]. This decision-making mechanism naturally leads to the emergence of weak and strong ties, weight heterogeneities, and a more realistic degree distribution in the integrated network structure [17]. Note that in this formulation c scales the strength of memory, which may capture the attitude being a *social keeper* or *social explorer* [21]. Here, for simplicity, we set $c = 1$ for each node and leave the exploration of the effect of varying memory strengths for future studies [40].

The second mechanism we consider is that of tie creation with *closure processes*. Here, *cyclic closure*, the tendency to form network triangles, shapes the social structure at mesoscopic scale, and is responsible for the emergence of communities [41]. The mechanism of *focal closure*, on the other hand, is independent of network structure and represents the formation of ties between individuals with shared attributes or interests. It is driven by the propensity to seek cognitive balance between connected egos [8, 42] as suggested by earlier theories in sociology [10, 43]. An applicative definition of *cyclic* and *focal closure* in general is given by Kumpula *et al.* [12, 34], who model cyclic closure as biased local search. To the contrary, *focal closure* is modelled as an unbiased global random search.

Finally, the third ingredient of the model is the process of *node removal*. It allows the model network to reach an equilibrium state where its overall structural characteristics become invariant of time.

B. Model definition

Here we introduce a time-varying network model, which integrates memory and reinforcement processes with closure mechanisms in the activity-driven model definition. Our ultimate goal is to provide a temporal network model with adjustable community structure and weight-topology correlations to understand their role in shaping the emergent network structure and co-evolving dynamical processes.

Our model takes N initially disconnected nodes and updates them in random order at each iteration step. In one iteration step, a node may become active with probability $a_i = \eta x_i$ or be deleted with probability p_d. Note that to keep the network size constant, for each deleted node, we add a new node to the network in the next iteration step. While a node i is alive, it remembers

its already connected neighbours $j \in V_t^i$ and the weight w_{ij}^t of interactions with each of them.

If the node i becomes active, then, depending on its actual degree n_i, it can attempt to form a new link with probability $p(n_i) = 1/(n_i + 1)$, or otherwise interact via an existing link with probability $(1 - p(n_i))$. In the latter case it randomly selects one of its neighbours j with probability $p_{ij}^w = w_{ij}^t / \sum_{k \in V_t^i} w_{ik}^t$ weighted by the number of their interactions. The two nodes then interact and increase their link weight w_{ij}^t by δ (*reinforcement process*).

On the other hand, if the node decides to form a new link, it may follow different scenarios. In all cases, the new tie will initially have unit weight $w^t = 1$. If the degree of the focal node is 0, it randomly picks another node from the entire network j (*focal closure*) and forms a tie. Otherwise, it attempts to create a new link with the triadic closure mechanism. First, it chooses one of its neighbours j randomly with a weighted probability p_{ij}^w. If j has no other neighbours than i, node i looks for another random node to interact with (*focal closure*) and forms a link. Otherwise, it looks for a random neighbour k of j ($i \neq k$) with a weighted probability p_{jk}^w. If k is not an already existing neighbour of i ($k \notin V_t^i$), the two nodes interact with probability p_Δ, and close the triad by forming a link. Otherwise, with probability $1 - p_\Delta$, node i follows the *focal closure* mechanism and instead forms a link with a randomly selected node (other than j and k). Finally, if k is already a neighbour of i, that is, $k \in V_t^i$, the two nodes interact and increase the weight of their existing link by δ (*reinforcement process*). At the end of each iteration step, all nodes finish their active interactions but remember their weighted egocentric network. For a pseudocode version of the algorithm, see Appendix A.

III. MODEL NETWORK ANALYSIS

In addition to the activity-driven model parameters whose values are fixed, our model has three intensive parameters, p_Δ, p_d, and δ. By varying these parameters, one can simulate a rich variety of time-varying networks with several emergent structural properties and correlations. In the following, we explore how the properties of the emerging network structure depend on time and on the intensive parameters, and whether those properties match with those of an empirical temporal network of mobile phone communications. The dataset we use here contains $633,986,311$ time-stamped mobile call communication records of $6,243,322$ customers of a mobile operator (market share $\approx 20\%$) in a European country. Customers are represented as network nodes, connected via $16,783,865$ weighted mutual links with weights defined as the number of calls between customers who mutually called each other at least once during the observation period. The presented network properties of the mobile phone call (MPC) network were calculated from a static,

aggregated representation of the social network structure obtained by integrating the temporal interactions over 6 months.

In the following, model networks were generated via large-scale numerical simulations with $N = 10,000$ nodes, and results were averaged over 100 independent realisations (if not noted otherwise). For each realisation, we measured the network parameters by considering links that are actually present in the network, *i.e.* we disregarded links of removed nodes.

A. Temporal features

The introduced network model is inherently temporal and simulates time-varying interactions between individuals. To explore its overall temporal behaviour, we measured two general network properties as a function of time. The first property is the average degree defined as $\langle k \rangle(t) = N^{-1} \sum_i k_i(t)$, where the sum runs over all nodes i, and k_i denotes the number of established connections of node i at time t (i.e. $k_i(t) = |V_t^i|$). Second, we measure the average local clustering coefficient $C(t)$, defined as the fraction of the real and possible numbers of triangles around a node given its number of links, averaged over the whole network [44]. C quantifies the density of triangles in a network; in social networks, the existence of communities typically gives rise to high triangle densities.
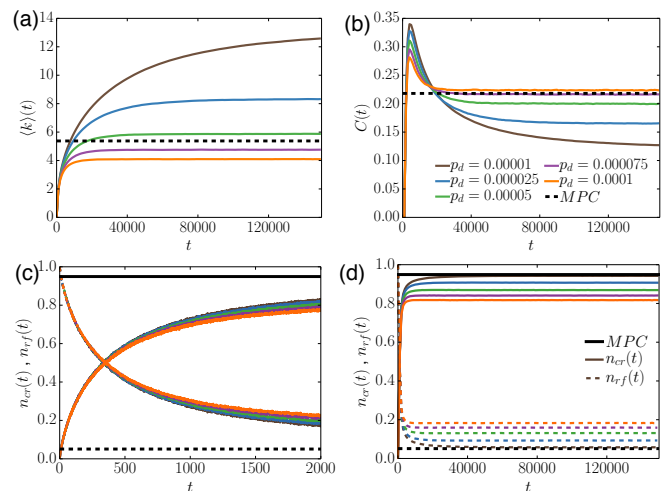

FIG. 1: Average network properties as the function of time. We depict the evolution of a) the average degree $\langle k \rangle(t)$, b) clustering coefficient C, and c,d) the fraction of events creating ($n_{cr}(t)$) or reinforcing ($n_{rf}(t)$) a link with various values of the deletion probability p_d (for exact values, see legend). For each calculation, we fixed $p_\Delta = 0.5$ and $\delta = 1$.

As follows from the model definition, the process starts from a set of disconnected agents, thus the measured properties are trivially zero at time $t = 0$. However, as time goes by and ties are formed via temporal interactions, $\langle k \rangle(t)$ starts increasing until the network reaches a
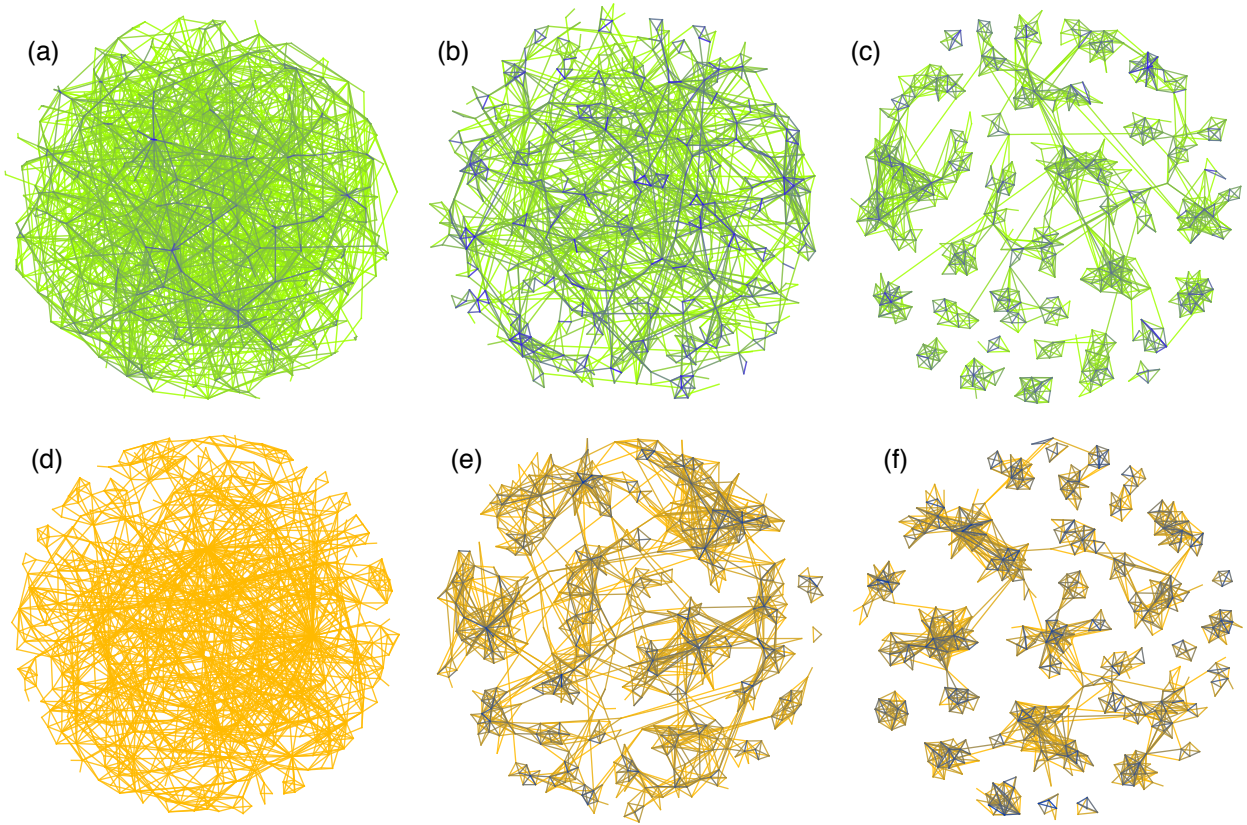
FIG. 2: Demonstration of the emerging structure in the time-varying network model. Panels (a-c) depict simulated networks with fixed $\delta = 1$ and varying $p_\Delta = 0.5$, 0.9, and 0.995 ($p_d = 4e-5$, $2e-5$, and $1.04e-5$) respectively. Panels (d-f) shows networks with fixed $p_\Delta = 0.995$ with varying $\delta = 0$, 0.5, and 1.5 ($p_d = 3.5e-5$, $1.7e-5$, and $8e-6$) respectively. Each panel depicts the actual structure of a network with $N = 500$, in its stationary state. Links are coloured according to their weight (darker link colour = stronger link weight).

stationary state with constant average degree. The time it takes to reach this equilibrium state strongly depends on the choice of the node deletion probability p_d as shown in Fig.1.a. If p_d is too small, nodes remain in the system for a long time and even nodes with small activity levels have time to evolve their egocentric network resulting in a slow relaxation to the stationary state. On the other hand, as p_d is increased due to the finite lifetime of nodes, less active nodes are removed before their egocentric structures are fully evolved. Because of this, the network reaches equilibrium faster. This explains the decrease of stationary average degree with increasing p_d. The average degree can be tuned approximate the empirical average degree of the MPC network (dashed line in Fig.1.a)

Measurement of the time dependence of the the average clustering coefficient (see Fig.1.b) yields a fairly similar scenario. However, here, after the initial increase of $C(t)$ it reaches a maximum, followed by a decrease and relaxation to a stationary value. This is because triangles tend to emerge right in the beginning of the process, followed by the evolution of strong ties. After nodes have created their first links and closed triangles

between them, interactions begin to frequently take place on existing links that are reinforced in the process. Thus the local search of active nodes is biased towards strong ties that are part of already formed triangles. In late times nodes attempt less frequently to create new triangles through the weak links that emerge throughout the dynamics by focal closure. Again, there is saturation due to node removal influenced by the rarely active but surviving nodes, who keep introducing random links in the network. By choosing an appropriate parameter values, the clustering coefficient can again be tuned to comparable to its corresponding empirical value (dashed line in Fig.1.b).

The above picture is fully supported by measurements capturing the fraction of events which create new links $n_{cr}(t) = \#e_{cr}/(\#e_{cr}(t) + \#e_{rf}(t))$, or reinforce already existing ones $n_{rf}(t) = \#e_{rf}/(\#e_{cr}(t) + \#e_{rf}(t))$. Here $\#e_{cr}$ ($\#e_{fr}$) denotes the number of such events at time t. As seen in Fig.1.c these measures take values of 1 and 0 (respectively) at time $t = 0$, as all events initially create new links. However they rapidly approach a constant value as it is visible in Fig.1.d. On the other hand, as p_d is increased, the number of events reinforcing already

existing links is increased, while events responsible for the creation of new links become less frequent. This is in accordance with our interpretation above. The measures can be tuned to reflect empirical network values here as well (black lines).

In the following, if not noted otherwise, we set $p_d = 0.5 \times 10^{-5}$ and run the simulations for $t = 150,000$ time steps to ensure the model networks to always reach their stationary state with average properties invariant in time.

B. Static features

Next we concentrate on the effect of different mechanisms on the emerging network structure in the stationary state. In Fig.2, we visually demonstrate the role of the cyclic closure (a-c) and link weight reinforcement (d-f) mechanisms in the emerging model networks. In panels Fig.2(a-c), we have kept the reinforcement increment constant, $\delta = 1$, while varying the cyclic closure probability p_Δ between 0.5 and 0.995. The node deletion probability p_d has been chosen to yield networks with suitable link density for visualisation (for exact parameters see the figure caption). When p_Δ is small, the emerging network structure is densely connected and appears more like a random structure since link creation is driven by focal closure. Nevertheless, because of the reinforcement process, weight heterogeneities emerge already in this case. More interestingly, by increasing p_Δ, communities are seen to emerge, with weight-topology correlations that are in line with the Granovetterian picture [8, 9], where strong ties connect nodes inside communities, while weak ties emerge between them (darker link colour = stronger link weight). However, focal closure alone is not sufficient for the emergence of community structure as shown in panels Fig.2(d-f). Here, the triangle formation probability has been kept constant, $p_\Delta = 0.995$, while the reinforcement parameter δ has been varied between 0 and 1.5. Even though a triangle would almost always be closed by the local search if a suitable node were found, without link weight reinforcement (d), the local search will be disperse and often resulting in focal closure because hitting a neighbour with no other neighbours. However, by increasing δ the local search will get more biased towards emerging strong ties and lead to the evolution of tight community structure. Note that since δ controls the strength of the local search mechanism, it also scales the size of the communities.

To quantitatively explore the emerging network structure as the function of the intensive parameters, we have measured the behaviour of general network properties. Fig.3 shows how the average degree of the model networks depends on p_Δ and δ (panels a and b, respectively) for different deletion probabilities p_d. The average degree of the empirical MPC network is displayed for comparison (dashed horizontal line). In each case, the average degree decreases when the parameters controlling the probability of cyclic closure p_Δ and the amount of

link reinforcement δ are increased, indicating that these parameters together with p_d strongly control $\langle k \rangle$. However, while the dependence of $\langle k \rangle$ on p_Δ is concave, $\langle k \rangle(\delta)$ is a convex function for any p_d. Moving beyond averages, panels c) and d) display the behaviour of the degree distributions $P(k)$ for various p_Δ and δ, indicating that there are strong degree heterogeneities. While the average degrees strongly depend on these parameters, the overall distribution shapes do not change much. The distribution of the empirical MPC is again displayed for reference, showing a very similar shape than those produced by the model.
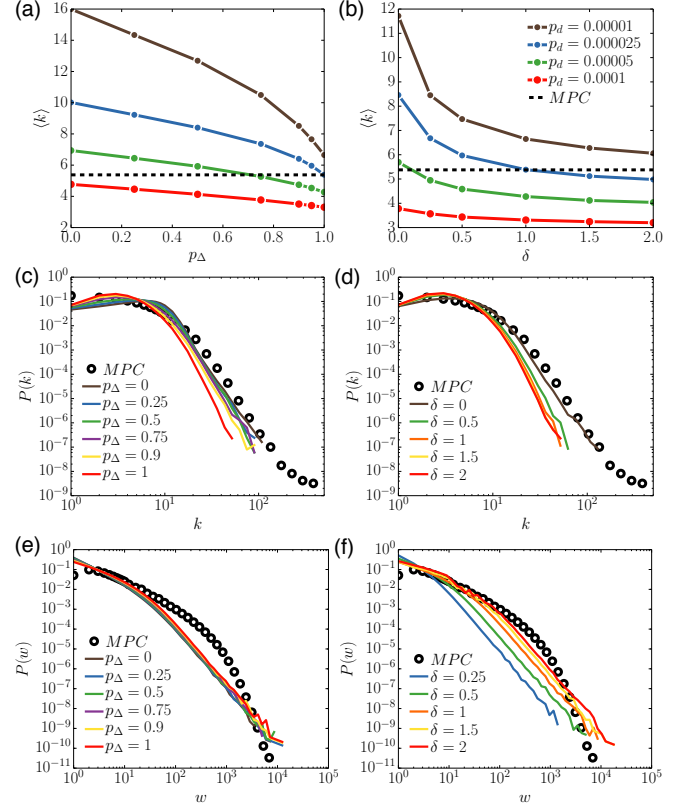

FIG. 3: General network measures as a function of the intensive parameters. (a) and (b) show the dependence of the average degree $\langle k \rangle$ on p_Δ and δ, respectively (for the former, $\delta = 1$ and the latter, $p_\Delta = 1$). Various values of p_d have been used (see legend in (b)). The dashed black line depicts the average degree $\langle k \rangle$ of the empirical MPC network. (c) and (d) depicts the degree distributions $P(k)$ of model networks with varying p_Δ and δ, while (e) and (f) show the corresponding weight distributions $P(w)$. In panels (c) and (e) ((d) and (f)), we kept $\delta = 1$ ($p_\Delta = 1$) and $p_d = 5 \times 10^{-5}$. Black circles denote the corresponding MPC distributions.

The strength of social ties, measured here as the number of dyadic interactions, is commonly observed to be distributed heterogeneously in social networks. This is also the case with our empirical MPC network (black circles in Fig.3 e and f). Tie strength heterogeneity is an emergent property of our model networks where it comes

from the preference of nodes to reinforce existing links. This mechanism is independent from the search strategies, as evident in Fig.3.e, where the model weight distributions are invariant to the selection of the cyclic closure probability p_Δ (δ and p_d are kept constant here). The emerging weight heterogeneities are the consequence of the intrinsic memory process, which inclines nodes to reinforce existing links rather than create new ones. Therefore, the weight heterogeneity naturally depends on the reinforcement parameter δ (see Fig.3.f), which in turn mainly scales the tail of $P(w)$ but does not affect not its functional shape, which appears to be a power-law with a constant exponent.

C. Higher-order correlations

Other than the observed degree and weight heterogeneities in the model networks, certain parameter ranges result in the emergence of rich community structure. In real social networks, communities can be characterised by higher-order correlations as they are built of closed triangles with unevenly distributed links of various strength in the structure. Stronger ties that are maintained by frequent interactions tend to connect nodes inside communities and shape the structure locally, while weak links with infrequent interactions are situated between communities and keep the social structure globally connected. To show whether the modelled temporal networks can recover these characteristics, we have performed three sets of measures and compared our findings to the empirical results.

As we have seen earlier, both the values of p_Δ and δ influence the community structure. However, we expect that p_Δ plays a dominant role here as it controls the triangular closing mechanism in the network. If $p_\Delta = 0$, links are created randomly and the clustering coefficient is very small as seen in Fig.4.a, where a constant $\delta = 1$ were set for each measurement. By increasing p_Δ, cyclic closure becomes more dominant, reflected in an increasing C with very high values as p_Δ goes to its maximum. The clustering coefficient of the MPC network is depicted as the horizontal dashed line. There are several sets of parameter values for which the model yields networks with similar clustering coefficient values, e.g. $p_\Delta \simeq 0.5$, $\delta \simeq 1$, and $p_d \simeq 5 \times 10^{-5}$ (see the crossing points of the black dashed and green solid lines in Fig.4.a). Not surprisingly, δ plays a weaker role in the emergence of triangles. The reinforcement mechanism introduces a bias in the local search for new ties, which leads to tighter communities and more closed triangles as we increase δ, reflected by the slightly increasing C in Fig.4.b.

Second, we have measured weight-topology correlations to check if the model networks are constructed according to the Granovetterian weak-tie structure. In his seminal paper [8] Granovetter suggests that the fraction of common friends of connected individuals is positively correlated with their tie strength (i.e. their link weight
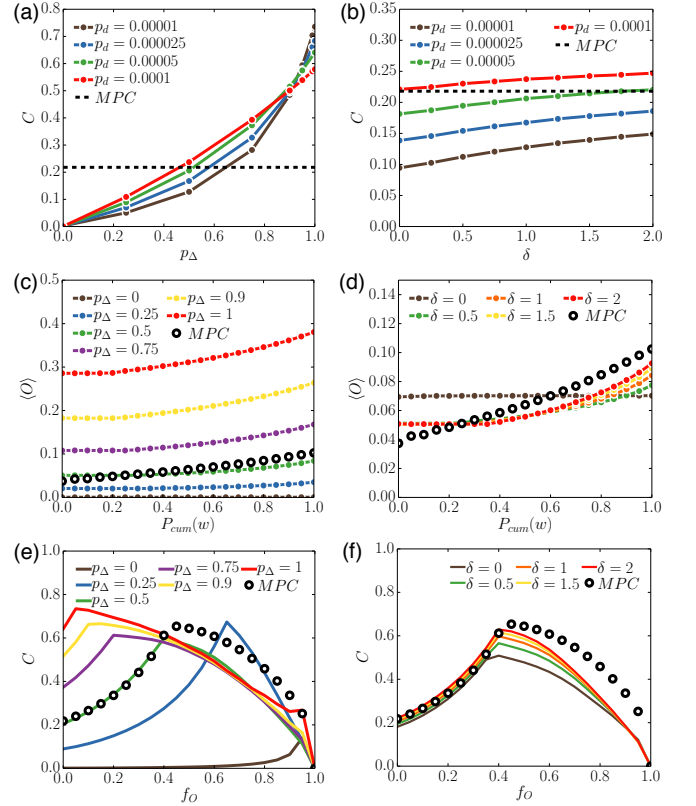

FIG. 4: Measures of higher-order correlations in the empirical and model networks. Panels (a) and (b) show the dependence of the average local clustering coefficient C on the parameters p_Δ and δ, respectively, for the empirical and model networks. Weight-topology correlations are shown as the function of the same parameters ((c) and (d), respectively) by measuring the average overlap $\langle O \rangle$ as the function of cumulative tie strength $P_{sum}(w)$. Panels (e) and (f) depict the average local clustering coefficient of the residual networks after removing a fraction f_O of nodes ranked by their overlap. Empirical results are shown in each panel with black dashed lines or circles, while measures on model networks are depicted by coloured solid lines with the corresponding parameters in the legend. In each measurements in panels (a,c,e) we kept $\delta = 1.0$, in panels (b,d,f) $p_\Delta = 0.5$, while in panels (c-f) $p_d = 5 \times 10^{-5}$.

here). The fraction of common friends can be measured by the link overlap [9] defined as $O_{ij} = n_{ij}/((k_i - 1) + (k_j - 1) - n_{ij})$ where n_{ij} is the number of common neighbours of nodes i and j, and k_i and k_j are their degrees. To quantify weight-topology correlations, we measure the average link overlap $\langle O \rangle$ as a function of cumulative tie strength $P_{cum}(w)$ that measures the fraction of links with tie strength smaller than w [9]. In the MPC network, this function reflects positive correlations between overlap and tie strength (black circles in Fig.4.c and d) in accordance with earlier observations [9]. If the model networks have their triangle-formation mechanism or weight reinforcement turned off ($p_\Delta = 0$ or $\delta = 0$), no such correlation is found. However, for any positive

values of p_Δ and δ, a positive weight-topology correlation emerges in accordance with the empirical observation and the hypothesis of Granovetter. For larger p_Δ (with constant $\delta = 1$), the function is shifted as more triangles evolve, which is increasing the average overlap (see Fig.4.c), while by increasing δ (and constant $p_\Delta = 0.5$) (see Fig.4.d) the correlation becomes stronger as ties with larger strength appear in the networks.

At last, we have checked how the clustering coefficient C changes by removing a fraction of links f_O in increasing order of overlap. The resulting function $C(f_O)$ can be divided into two regimes. Trivially the removal of links with $O = 0$, which connect communities, does not decrease the number of triangles in the network but only decreases the degrees of nodes by removing links not participating in triangles. Consequently, by the removal of these links the clustering coefficient must increase. On the other hand, the removal of links with $O > 0$ that connect nodes inside communities reduces the number of triangles, resulting a decrease in C. This behaviour is observed in the empirical system (black circles in Fig.4.e and f) and also for the model networks. In Fig.4.e as $p_\Delta \to 0$ no triangles evolve in the network, and thus most of the links have $O = 0$, and the first regime is extended. In the other extreme, as $p_\Delta \to 1$, most of the links have non-zero overlap and the second regime dominates. The best match with the empirical curve appears when $p_\Delta \simeq 0.5$ (red curve in Fig.4.e) and $\sim 40\%$ of links evolve with $O = 0$. Note that the second regime of this function is very sensitive to the fine-grained structure of the actual communities, which causes the difference between the model and empirical curves. In addition, by keeping $p_\Delta = 0.5$, we have checked the δ dependence of this function (see Fig.4.f). As the tie reinforcement δ increases, links form tighter communities and the clustering coefficient increases in this case.

IV. EFFECTS ON SPREADING PROCESSES

We conclude our modelling study by demonstrating its capacity in simulating dynamical processes co-evolving with the temporal network structure with scalable integrated social mechanisms. For this reason, we have selected the simplest possible spreading process, the susceptible-infected (SI) model [45], where each node can be in one of two mutually exclusive states; susceptible (S) or infected (I). Initially, every node is in the state S, except for a random seed with state I. The infection passes from any node in the I state to any node in the S state via a temporal interaction, but independent of the direction of the actual temporal link [22]. In order to assure that the network is in the stationary state we initiated the spreading process after the temporal network of size $N = 10,000$ has evolved for $t = 50,000$ iterations. After this time we measured the $i(t) = I(t)/N$ fraction of infected nodes until $t = 150,000$, where the process has reached maximal penetration (see Fig.5.a and b). Note

that maximal penetration is not converging to 1 as there are always new susceptible nodes introduced in the network, moreover the temporal structure is not necessarily connected. For simplicity, we set the probability of transmission per interaction event to $\lambda = 1$ (note that then the speed of transmission between two nodes is only limited by the frequency of their interactions). Further, we consider interactions as undirected and therefore they may transmit the interaction both ways.
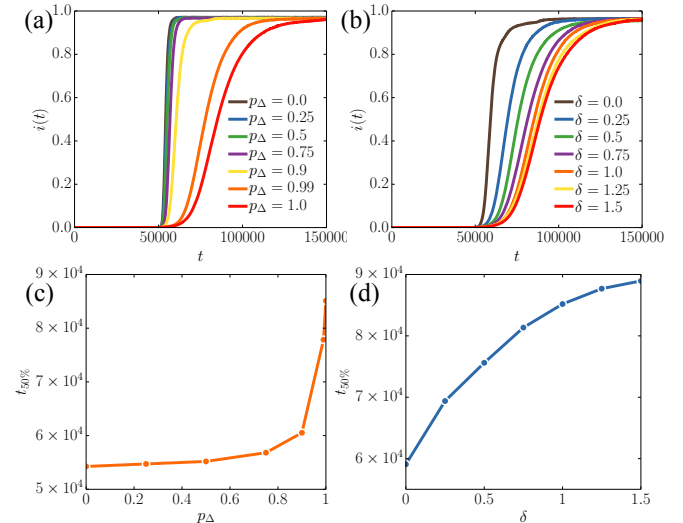

FIG. 5: Susceptible-Infected process co-evolving with the time-varying network model (a, b) Fraction of infected nodes as a function of time for different values of p_Δ and δ (for exact parameters see legend). (c, d) Time $t_{50\%}$ of 50% penetration for different values of p_Δ and δ. While in case of panels (a, c) we have kept $\delta = 1$, for panels (b, d) we have used $p_\Delta = 0.99$. Each simulation was run for $t = 150,000$ iteration steps with $p_d = t \times 10^{-5}$ on networks of size $N = 10,000$. Results were averaged over 100 realisations, where we initiated the process from a single random seed at $t = 50,000$.

The simulation results depicted in Fig.5.a and b show that both the spreading process becomes slows when either the cyclic closure probability p_Δ or the weight reinforcement parameter δ are increased. While in the first case this is mainly due to the emerging tight communities, in the second case weight-topology correlations also play a role. The same conclusions can be drawn by measuring the speed of spreading, quantified as the time $t_{50\%}$ when the process has reached 50% penetration. Tight communities evolve as $p_\Delta \to 1$, which constrains the spreading process and explains the convex shape of the curve in Fig.5.c. On the other hand, weight-topology correlations appear even for small values of δ (as we have shown in Fig.4.c). These immediately slow down the speed of spreading and cause the concave shape of the function in Fig.5.d. Thus, one can conclude that even though more pronounced community structure has considerable effects on the spreading process, already weak weight-topology correlations can strongly affects

the speed of spreading.

V. CONCLUSIONS

The temporal network representation takes into account the time-varying nature of interactions between entities instead of considering only the static network structure of their connections [14]. So far, this approach has mainly focussed on analyzing empirical data, while there is still a lack of appropriate network models. In this paper we introduce a model of temporal social networks with emerging heterogenous structure, communities, and higher-order correlations. As a starting point, we take the recently introduced modelling framework of activity-driven networks with heterogeneously distributed activities of individuals [29] and memory processes [17]. We extend this model with three mechanisms that have been earlier used for generating static weighted networks [12], such as *social reinforcement*, *cyclic* and *focal closure*. These mechanisms arguably drive the tie formation of individuals and are responsible for the emergence of communities and the global connectedness of the network. In addition, we introduce a *node removal* process allowing the model network to reach an stationary state, where its overall structural characteristics become invariant of time. By adjusting the model parameters, it is possible to recover several characteristics of a real-world temporal network of mobile phone calls. These include heterogeneous degrees and weights, rich community structure, and weight-topology correlations and weak tie properties as suggested by Granovetter [8].

The main advantages of this model are that (a) it is able to mimic the time-varying nature of interactions characterising several real systems; (b) it helps to understand the importance of different mechanisms in shaping the emerging network structure; and (c) it allows us to test their effect in a scalable fashion on co-evolving dynamical processes. This last point is of special importance, as *e.g.* spreading processes have been shown to have significantly different critical behaviour when co-evolving with time varying interactions. In addition, by varying the model parameters one can control the average degree, strength of ties, size of communities, clustering, and interconnectedness of the network. These features can control e.g. the speed of spreading as we demonstrated via a simple study on SI processes, but play crucial effects in case of other processes like random walks, the diffusion of information, epidemics, or social contagion processes, making the model an ideal testbed for simulating such processes.

Appendix A: Algorithmic model definition

Here we describe the algorithmic implementation of the temporal network model introduced in the main text. The main function **Temporal network model(**$G_{t=0}$, T, p_Δ, p_d, δ**)** takes as input a weighted network with empty link set $G_{t=0} = (V, \emptyset, \{a_i\}, \{w_{ij}^t\})$ with size N, pre-assigned node activity probabilities per unit time $\{a_i\}$, and $initial link weights \{w_{ij}^{t=0} = 0\}$. Further parameters are T: the number of iterations; p_Δ: probability of triadic closure; p_d: probability of node deletion; and δ: link reinforcement increment. In the pseudo code, $n_i = |V_t^i|$ denotes the current degree of node i; $rand()$ is a pseudo-random number generator, which returns a rational number between $[0, 1]$; and $p_{ij}^w = w_{ij}^t / \sum_{k \in V_t^i} w_{ij}^t$ is the probability to select a random neighbour j from the current neighbour set V_t^i of node i weighted by the number of interactions between them performed since the link was created up to time t ($\sum_{j \in V_t^i} p_{ij}^w = 1$). Temporal events between nodes i and j performed at the current iteration step t are denoted by $[t, i, j]$.

ALGORITHM 1: *Temporal network model*

Input : $G_{t=0}$, T, p_Δ, p_d, δ

for $t = 1$ to T **do**
  for N randomly selected node $i \in V$ **do**
    if $rand() \le p_d$ **then**
      remove all links of node i ▷ *delete and re-insert i*
    end if
    if $rand() \le a_i$ **then**   ▷ *node i is active*
      if $rand() \le p(n_i)$ **then**   ▷ *link creation*
        if $n_i = 0$ **then**   ▷ *node i has no neighbour*
          Focal closure$(i, \{i\}, G_t)$
        else
          Cyclic closure(i, G_t)
        end if
      else   ▷ *link reinforcement*
        Select a j neighbour of i with probability p_{ij}^w
        Event $[t, i, j]$
        $w_{ij}^t + = \delta$
      end if
    end if
  end for
end for

The main function calls two subroutines called **Cyclic closure**(i, G_t) and **Focal closure**(i, G_t, X) with input parameters i denoting the currently active node; G_t the current weighted network structure; and X exception set of nodes to connect in the current call.

ALGORITHM 2: *Cyclic closure*: as input i is the current active node; and G_t is the current weighted network

---

Input : i, G_t

Select a j neighbour of i with probability p_{ij}^w
if $n_j = 1$ **then**   ▷ *node j has one neighbour*
  ***Focal closure**(j, $\{i,j\} \cup V_t^i$, G_t)*
else
  Select a $k(\neq i)$ neighbour of j with probability p_{jk}^w
  if $(i,k) \notin E_t$ **then**   ▷ *triangle (i,j,k) is not closed*
    if $rand() < p_\Delta$ **then**
      Event $[t,i,k]$
      $w_{ik}^t = 1$
    else
      ***Focal closure**(j, $\{i,j,k\} \cup V_t^i$, G_t)*
    end if
  else
    Event $[t,i,k]$
    $w_{ik}^t += \delta$
  end if
end if

ALGORITHM 3: *Focal closure*: as input i is the current active node; X is the exception set of nodes to connect; and G_t is the current weighted network

---

Input : i, X, G_t

Select a random node $j \in V \setminus X$
Event $[t,i,j]$
$w_{ij}^t = 1$

---

[1] C. Castellano, S. Fortunato, and V. Loreto, Rev. Mod. Phys **81**, 591 (2009)

[2] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences)*, 1st ed., **8** (Cambridge University Press, 1994).

[3] D. Lazer, et.al., Science **323**, 721 (2009)

[4] A. Vespignani, Science **325**, 425 (2009)

[5] D. J. Watts, Proc. Natl. Acad. Sci. **99**, 5766 (2002)

[6] D. J. de Solla Price, Science **149**, 510 (1965)

[7] R. Albert and A.-L. Barabási, Rev. Mod. Phys **74**, 47 (2001)

[8] M. S. Granovetter, Am. J. Sociol. **78**, 1360 (1973)

[9] J. P. Onnela, et. al., Proc. Natl. Acad. Sci. **104**, 7332 (2007)

[10] A. Rapoport, B. Math. Biophys. **19**, 257 (1957)

[11] S. Fortunato, Phys. Rep. **486**, 75 (2010)

[12] J. M. Kumpula, et. al., Phys. Rev. Lett. **99**, 228701 (2007)

[13] G. Palla, A.-L. Barabási, and T. Vicsek, Nature **446**, 664 (2007)

[14] P. Holme and J. Saramäki, Phys. Rep. **519**, 97 (2012)

[15] L. Kovanen, et. al., *Temporal Networks, Understanding Complex Systems*, eds. P. Holme and J. Saramäki (Springer Berlin Heidelberg, 2013) pp. 119133

[16] K. Zhao, J. Stehlé, G. Bianconi, and A. Barrat, Phys. Rev. E **83**, 056109+ (2011)

[17] M. Karsai, N. Perra, and A. Vespignani, Sci. Rep. **4**, 4001 (2014)

[18] A.-L. Barabási, Nature **435**, 207 (2005)

[19] A.-L. Barabási, *Bursts: The Hidden Pattern Behind Everything We Do*, 1st ed. (Dutton Adult, 2010).

[20] L. Kovanen, K. Kaski, J. Kertész, and J. Saramäki, Proc. Natl. Acad. Sci. **110**, 18070 (2013)

[21] G. Miritello, R. Lara, M. Cebrian, and E. Moro, Sci. Rep. **3**, 1950 (2013)

[22] M. Karsai, et. al., Phys. Rev. E **83**, 025102 (2011)

[23] M. Kivelä, et. al., J. Stat. Mech., P03005 (2012)

[24] G. Miritello, E. Moro, and R. Lara, Phys. Rev. E **83**, 045102+ (2011)

[25] R. Lambiotte, L. Tabourier, and J.-C. Delvenne, Eur. Phys. J. B **86**, 320 (2013)

[26] T. Takaguchi, N. Masuda, and P. Holme, PLoS ONE **8**, e68629+ (2013)

[27] B. Min, K. I. Goh, and A. Vazquez, Phys. Rev. E **83**, 036102 (2011)

[28] L. Rocha, A. Decuyper, and V. Blondel, Dynamics On and Of Complex Networks, Volume 2 , *Modeling and Simulation in Science, Engineering and Technology*, ed. by A. Mukherjee, et. al. (Springer New York, 2013) pp. 301314+.

[29] N. Perra, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, Sci. Rep. **2**, 469 (2012)

[30] M. V. Tomasello, et. al., Sci. Rep **4**, 5679 (2014)

[31] N. Perra, et. al., Phys. Rev. Lett. **109**, 238701 (2012)

[32] M. Starnini and R. P. Satorras, Phys. Rev. E **87**, 062807 (2013)

[33] S. Liu, N. Perra, M. Karsai, and A. Vespignani, Phys. Rev. Lett. **112**, 118702 (2014)

[34] J. M. Kumpula, et. al., Comp. Phys. Comm **280**, 517 (2009)

[35] C. Cattuto, et. al., PLoS ONE **5**, e11596+ (2010)

[36] A. Chmiel, K. Kowalska, and J. A. Holyst, Phys. Rev. E **80**, 066122+ (2009)

[37] J.-P. Onnela, et. al., New. J. Phys. **9**, 179 (2007)

[38] R. I. M. Dunbar, J. Hum. Evo. **22**, 469 (1992)

[39] J. Saramäki, et. al., Proc. Natl. Acad. Sci. **111**, 942 (2014)

[40] E. Ubaldi, et. al., to be bublished (2015)

[41] G. Bianconi, R. K. Darst, J. Iacovacci, and S. Fortunato, Phys. Rev. E **90**, 042806 (2014)

[42] F. Heider, J. Phsy. **21**, 107 (1946)

[43] K. H. Wolff, *The Sociology of Georg Simmel* (Nabu Press, 2011).

[44] M. Newman, *Networks: An Introduction*, 1st ed. (Oxford University Press, USA, 2010).

[45] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks*, 1st ed. (Cambridge University Press, 2008).